# ISOTHERMAL STRAND DISPLACEMENT AMPLIFICATION APPLICATIONS FOR HIGH-THROUGHPUT GENOMICS

John C. Detter[1], Jamie M. Jett[1], Andre R. Arellano[1], Alicia R. Ferguson[1], Mei Wang[1], John R. Nelson[2], Susan M. Lucas[1], Jarrod Chapman[1], Yunian Lou[1], Dan Rokhsar[1], Paul M. Richardson[1]*, and Trevor L. Hawkins[1]

[1] US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598.

[2] Amersham Biosciences, 800 Centential Avenue. PO Box 1327 Piscataway, NJ 08855.

* Corresponding Author

**Amplification of source DNA is a nearly universal requirement for molecular biology applications. The primary methods currently available to researchers are limited to *in vivo* amplification in *E. coli.* hosts and the Polymerase Chain Reaction. Rolling-circle DNA replication is a well-known method for synthesis of phage genomes and recently has been applied as rolling circle amplification (RCA) of specific target sequences[1] as well as circular vectors used in cloning[2]. Here, we demonstrate that RCA using random hexamer primers with Φ29 DNA polymerase can be used effectively in amplification of DNA for sequencing plasmids, cosmids and BACs. This procedure can also be used for the production of DNA from cosmids and BACs for use in sub-cloning and subsequent shotgun sequencing. In addition, we show that whole bacterial genomes can be effectively amplified from cells or small amounts of purified genomic DNA without apparent bias for use in downstream applications including whole genome shotgun sequencing.**

Rolling circle replication is typically employed by certain viruses to replicate their circular genomes in nature[3]. Rolling circle amplification (RCA) has traditionally been used in the laboratory using specific primer sequences to replicate circular sequences for

the generation of tandem repeat libraries and as reporter constructs[1,4]. Rolling-circle amplification has recently been employed using exonuclease resistant random hexamer primers and Φ29 DNA polymerase to generate large quantities of DNA from traditional vectors in an isothermal reaction by way of multiply-primed template and DNA strand displacement[2]. This method allows amplification of circular input DNA up to a reported 10,000-fold level. At the US Department of Energy (DOE) Joint Genome Institute (JGI) we are interested in methods and approaches to expedite many of our DNA manipulation processes. Using a commercially available DNA sequencing template amplification kit (TempliPhi™, Amersham Biosciences), we have been able to utilize strand displacement DNA amplification and RCA for a number of important steps.

The first application of RCA technology to be used at the Production Genomics Facility (PGF) was to replace our existing plasmid DNA isolation methods[5,6]. We began by optimizing RCA conditions for robust amplification and sequencing of plasmids and electrophoresis on capillary instruments including the MegaBACE 1000™ and the ABI 3700™ sequencers. We have found that having a uniform DNA template concentration significantly increases success rates for capillary sequencing. As shown in Figure 1A, amplification of plasmids from saturated bacterial glycerol stocks produces similar quantities of DNA in each well. The PGF has now completely transitioned from plasmid purification of overnight growths to strand displacement amplification for template generation. Other advantages of this technology include the complete elimination of labor intensive steps of bacterial lysis and DNA purification as well as streamlining of the production process from 10 steps down to 4. In addition, we were able to move this part of the process into 384 well format from the previously used 96 well approach, with prospect for even higher density and alternative formats. More importantly, the production results from this RCA approach have been extremely encouraging, yielding over 2.4 billion high quality raw bases since being phased into production. Over this time, average pass rates for this process have been 92% with read lengths greater than 600 bases per lane, allowing for a 75% increase in Phred 20 bases output per day compared to our original platform[5,6] (Fig. 1B).

We have demonstrated that this process also can be used to sequence plasmids carried by yeast cells. Plasmids (pGADT7) were transformed into yeast cells (Y187) for the purpose of a yeast two-hybrid assay (CLONTECH). Yeast cultures scoring positive in these assays usually must undergo days of growth and rather involved procedures for isolating plasmid DNA for sequencing. We have shown that a simple lysis procedure like that used for amplification of plasmids from *E. coli* also can be used for amplification and sequencing of plasmids from yeast colonies or cultures (data not shown).

A further application of the RCA process has been in the amplification of cosmids and bacterial artificial chromosomes (BACs) for end sequencing. This is an important step in our genomic sequencing strategy since knowing the sequence at the cosmid and BAC insert ends greatly improves the long range contiguity of sequence assemblies, especially in whole genome shotgun approaches[7,8]. Our standard approach to sequencing large insert, low copy number clones was to use methods such as Qiagen® Maxi prep[9] or TIGR BAC isolation prep[10] which typically yields 700-1000 ng of DNA and, in our hands, approximately 400 Phred 20 bases with an average pass rate of 60%. With the aim of reducing labor and time we began with an approach similar to our plasmid RCA methods where we amplify directly from glycerol stock culture for direct end sequencing. However, unlike the results obtained with smaller molecular weight plasmids, amplification of cosmids and BACs directly from *E. coli* cultures seems to result in higher amounts of background *E. coli* amplification relative to clone amplification. This can result in erratic sequencing results. Based on these results we performed a simplified single 96 well TIGR prep[10] followed by a 20 hour RCA in an attempt to overcome the high *E. coli* background amplification. With this approach we were able to achieve a 70% pass rate in a 96 well format with an average read length of 450 bp. This abbreviated TIGR approach mixed with RCA reduced labor by an estimated 50% while increasing our pass rate, readlength and throughput. Additional experiments are currently underway to further reduce *E. coli* chromosomal DNA amplification and optimize the process for larger insert constructs.

We have extended the use of this technology to amplify cosmids and BACs to produce DNA for use in library construction. A typical shotgun library is constructed by shearing large insert clones (cosmids, P1s, BACs) or genomic DNA and subcloning a specific isolated fragment size into M13 or pUC type vectors. This usually involves the use of large culture volume DNA isolation methods such as Qiagen® Maxi prep[9]. We wished to explore the use of RCA technology to improve the efficiency of this process. To this end, we amplified four previously sequenced BACs directly from frozen glycerol stocks using strand displacement and random hexamer primed amplification (TempliPhi™), and then used this DNA in our standard library construction methods. While this process produced libraries that could be sequenced and used to assemble large inserts, the *E. coli* contamination was higher than that obtained with the traditional process. These results are in line with our end sequencing experiments, and were not totally unexpected since cosmids and BACs are very low copy number systems[11,12]. Additional experiments using specific vector primers are currently underway to overcome the *E. coli* bias.

Perhaps a more far reaching application of strand displacement amplification has been in the amplification of whole microbial genomes. Isolating DNA from microbes can be a laborious and low yielding process[13,14]. Furthermore, greater than 90% of microbial species are thought to be non-culturable. With this in mind, we set out to not only amplify microbes of varying genome size from both purified DNA and cell suspensions, but also to use the product as template for PCR, direct sequencing, and shotgun library construction.

As little as 50 ng of genomic DNA from five different microbes ranging in size from 1 to 4.6 Mb was amplified with the TempliPhi™ kit (Fig. 2A). Figure 2B shows a dilution series performed on *Xylella* cell suspensions. As few as 1000 cells were amplified to produce DNA that was detectable on an agarose gel. Thus, small amounts of DNA or small numbers of cells may serve as template for the random hexamer strand displacement amplification process.

Next we wanted to test the usefulness of the RCA DNA we produced. RCA product from *Xylella fastidiosa* cells was successfully used as template for PCR using

several dozen primer pairs spread randomly throughout the *Xylella* genome. Amplicon sizes of 1 kb to 4 kb were produced as expected when compared to the control amplicons produced from *Xylella* genomic DNA PCR (data not shown). The same RCA product used for PCR was also successfully directly sequenced using several *Xylella fastidiosa* specific primers (data not shown). In addition, the RCA produced *Xylella fastidiosa* DNA was sheared, blunt ended, size selected and ligated into a standard pUC18 cloning vector (Fig. 2C). A sample of colonies were PCR amplified for quality control of size and insert ratio demonstrating that ~ 96% of the colonies contained insert and that all inserts gave the expected 2.5-4 kb product (Fig 2C).

We wished to determine if the amplified DNA was representative of the *Xylella* genome and if the library made from that DNA had qualities similar to that of a library made from unamplified genomic DNA. The original *Xylella* library was made from 4 μg of genomic DNA prepared from *Xylella* cells using standard techniques. This library was sequenced to a depth of approximately 7x and assembled with Phrap[15]. Approximately 3000 subclones from the RCA amplified library were picked and sequenced. Trimmed reads from the RCA assembly were aligned with contigs greater than 10kb from the genomic assembly (Fig. 3). Coverage was computed as the number of bases that were covered by the reads over the sum of the lengths of the contigs. Coverage for the RCA library as well as a similar sample of reads from the genomic library is shown in Figure 3.

This experiment shows that a whole genome shotgun library can be produced from a small number of cells without isolating genomic DNA. Additionally, indications from sequence information derived from this library are that the amplified DNA is representative of the original genome, and that the coverage is what is expected from a library sequenced to this depth. Efforts are currently underway to decrease the amount of input DNA required for this procedure to as little as one cells contents of chromosomal DNA.

The use of random hexamer primed isothermal strand displacement amplification using Phi29 DNA polymerase offers many advantages over traditional techniques in a variety of molecular biology applications. We have taken advantage of commercially available reagents (TempliPhi™) to increase the efficiency of the production sequencing

process at the US DOE Joint Genome Institute. The plasmid amplification method has significant advantages over traditional template generation by greatly simplifying the process and will be a benefit for genome centers as well as core labs, service centers and academic researchers that isolate and sequence from a variety of vectors. This approach has few steps, can be highly automated and is already giving excellent results at our production genomics facility. The amplification of large clones for DNA sequencing and library construction using RCA also has significant process advantages. We think that the implementation of this approach will simplify large insert clone end sequencing, leading to reduced costs and increased use in conjunction with whole genome shotgun approaches. Its use in library construction will also lend significant advantages that we see leading to the automation of this part of the process.

The final and perhaps the most interesting application is the amplification of microbial genomic DNA. It is estimated that greater than 90% of all microbes are non culturable and many live in communities which are not easily separated. The use of Phi29 DNA polymerase and random hexamer primed isothermal strand displacement amplification as described here demonstrates that microbial DNA ranging up to 4.6 Mb in size can be easily amplified leading to a product that can be used for PCR, direct sequencing with specific primers, and random whole genome shotgun construction and sequencing. Eliminating the need for growing an organism in culture to obtain its genomic sequence opens the door to countless new genes. It is clear that this technology offers many advantages and process improvements to the genome sequencing community.

## Experimental Protocol

1. Template amplification: 2 μl aliquots of saturated E.coli cultures (DH10B or DH5α) containing pUC 18 vector with random 2-4 kb DNA inserts, Cosmid vector Lawrist 4 vector with random 40 - 60 kb DNA inserts, or BAC vector pBeloBAC with random 100-150 kb DNA inserts grown in LB/ glycerol (7.5%) were added to 8 μl of a 1 mM Tris-HCl pH 8 + 0.05 mM EDTA denaturation buffer. The mixtures were heat lysed

at 95°C for 5 min then placed @ 4C for 5 min. To these denatured products 10 µl of an RCA reaction mixture (TempliPhi™ DNA sequencing template amplification kit, Amersham Biosciences) were added. The amplification reactions were then carried out at 30°C for 12 hr. The reaction products were heat inactivated at 65°C for 10 min then placed at 4°C until used for sequencing.

2. Sequencing: Aliquots of the 20 µl amplified RCA products were sequenced with standard M13 –20 and –40 primers or primers specific to cosmid and BAC vectors. The reactions contained 1 ul RCA product, 4 pmoles primer, 5 µl $dH_2O$, and 4 µl DYEnamic™ ET terminator sequencing kit (Amersham Biosciences). Cycle sequencing conditions were 30 rounds of 95°C-25 sec., 50°C-10 sec., 60°C-2 min., hold at 4°C. The reactions were then ethanol precipitated or purified by a magnetic bead protocol (Elkin, submitted) and run on a MegaBACE 1000 (Amersham Biosciences) or an ABI 3700 (Applied Biosystems) capillary DNA sequencer.

3. Cosmid / BAC / Microbe DNA RCA amplification and subcloning: Cosmid and BAC saturated cultures were amplified as above. Five 20 µl reactions were amplified for each Cosmid or BAC and then mixed to a total volume of 100 ul for each sample. The amplified samples were then sheared in a HydroShear™ (Gene Machines) at speed code 11 for 25 cycles (gives an average band size of between 2 – 4 kb). 80 µl of each sample was then end-repaired using 15 µl of 10x klenow buffer, 21 µl of 10 mM dNTPs, 12 U of T4 DNA Polymerase, and 60 U of Klenow Fragment for 40 min at room temperature (RT), heat inactivate for 15 min. @ 70C, and place on ice for 5 min. Loading dye was then added to each sample. They were loaded onto an agarose gel and run until there was ~ 1 cm separation between the 2 and 4 kb size marker. An ~ 3 mm gel slice was taken in the 3 kb region for each sample. Each gel slice was purified using the QIAquick™ gel extraction kit (Qiagen). Samples were eluted in ~ 50 µl according to protocol. Using standard ligation procedures, ~ 200 ng of each sample was ligated into pUC18 using the Fast-Link™ DNA Ligation Kit (Epicentre Technologies). Colonies from each sample were picked, grown and plasmid DNA isolated for sequencing. Sequencing was carried out using both DYEnamic™ ET

terminator sequencing kit (Amersham Pharmacia Biotech) and BigDye™ terminator sequencing kit (Applied Biosystems) methods.

## Figure legends

**Figure 1. (A)** Strand displacement amplification of plasmids in production. 2 µl aliquots of saturated *E. coli* culture were heat lysed and amplified with TempliPhi™, Amersham Biosciences, for 12 hours at 30°C (see experimental protocol). 3 µl of product from 24 wells of a 384-well plate were run on a 1% agarose gel and stained with Ethidium Bromide. Lanes: (s) is a lambda mass standard of 250 ng, (1-24) amplified product, (m) lambda Hind III size marker with 23, 4.3, 2.3 kb marked. Note the level of amplification consistency across all 24 samples. This consistency allows for higher quality sequencing output. **(B)** Results of sequencing performed on amplified templates since introduction into our production environment. The most recent 3 months signify full implementation of TempliPhi™ into our production process. Samples were run on MegaBACE 1000 (Amersham Biosciences) capillary sequencers. Pass Rate (right Y-axis) is defined as the percentage of lanes per 96 well plate with a minimum of 50 bases with a Phred score $\geq$ 20. Read length (left Y-axis) is defined as the number of Phred $\geq$20 bases per passing lane. Current results show average pass rate to be > 93% with an average read length > 630 bp.

**Figure 2. (A)** Several microbial whole genome amplifications using the random hexamer Phi29 strand displacement amplification approach. 2 µl of cells resuspended in 1x PBS or 50 ng of DNA were heat lysed and amplified with TempliPhi™, Amersham Biosciences, for 12 hours at 30°C (see experimental protocol). 5 ul of each product were run on a 1% agarose gel and stained with Ethidium Bromide. Lanes: (m) lambda Hind III size marker with 23, 4.3, 2.3 kb marked and (1-8) various amplified products identified in the legend to the left. Within the legend each microbe is listed along with its template source and reported genome size. **(B)** Amplified *Xylella fastidiosa* (Dixon, Almond strain) cell dilutions. Various cell dilutions, listed in legend to the left, were heat lysed and amplified with TempliPhi™, Amersham Biosciences, for 12 hours at 30°C (see experimental protocol). 5 µl of each product were run on a 1% agarose gel and stained with Ethidium Bromide. Lanes: (A-F) amplified products, (m) lambda Hind III size marker with 23, 4.3, 2.3 kb marked, (s) lambda mass standards ranging from 125 ng to

500 ng. **(C)** *Xylella fastidiosa* cells amplified for library creation. 2 µl of *Xylella fastidiosa* cells were amplified, sheared, and subcloned (see experimental protocol). The red hashed box on the sheared product identifies the 2.5-4 kb band extracted to use as source material for library creation. To test the initial quality of the subclone library a PCR QC was performed on 24 clones using standard methods. Lanes: (m) lambda Hind III size marker with 23, 4.3, 2.3 kb marked and (1-24) PCR product. All PCR amplicons are within the expected 2.5-4 kb size range.

**Figure 3.** *Xylella* amplified & genomic content comparison. The original *Xylella* library made from made from unamplified DNA was assembled with Phrap (Phil Green and used as a reference. Approximately 3000 reads (0.5x) from both the amplified (red) and unamplified genomic (yellow) libraries were aligned with contigs from the 7x genomic library assembly. The largest contig assembled was 130 kb and is designated at the top with a red arrow going from left to right. Distribution of reads is similar for both libraries across the contig. Estimated genome size, which agrees with published data[16] and genome coverage are also similar for both libraries.

**This work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, under Contracts No. W-7405-Eng-48, No. DE-AC03-76SFOO098, and No. W-7405-ENG-36.**

## References


1. Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C. & Ward, D.C. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* **19**, 225-232 (1998).

2. Dean, F.B., Nelson, J.R., Giesler, T.L. & Lasken, R.S. Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification. *Genome Res.* **11**, 1095-1099 (2001).

3. Kornberg, A. & Baker, T.A. *DNA replication*. W.H. Freeman and Company, San Francisco. (1992).

4. Fire, A. & Xu, S.Q. Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci*. **92**, 4641-4645 (1995).

5. Hawkins, T.L., McKernan, K.J., Jacotot, L.B., MacKenzie, J.B., Richardson, P.M. & Lander, E.S. A magnetic attraction to high-throughput genomics. *Science.* **276**, 1887-1189 (1997).

6. Elkin, C.J. et. al. High-throughput plasmid purification for capillary sequencing. *Genome Research*. **11**, 1269-74 (2001).

7. Venter, J.C. et al. The sequence of the human genome. *Science.* **291**, 1304-1351 (2001).

8. McPherson, J.D. et al. A physical map of the human genome. *Nature.* **409**, 934-941 (2001).

9. Pham, T.T., Chillapagari, S. & Suarez, A.R. Preparation of pure plasmid or cosmid DNA using single-strand affinity matrix and gel-filtration spin columns. *Biotechniques*. **20**, 492-497 (1996).

10. Kelley, J.M., Field, C.E., Craven, M.B., Bocskai, D., Kim, U.J., Rounsley, S.D. & Adams, M.D. High throughput direct end sequencing of BAC clones. *Nucleic Acids Research. Mar.* **27**, 1539-1546 (1999).

11. Frame, R. & Bishop, J.O. The number of sex-factors per chromosome in Escherichia coli. *Biochemical Journal.* **121**, 93-103 (1971).

12. Sambrook, J. & Russell, D.W. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. (2001).

13. Jones, A.S. The isolation of bacterial nucleic acids using cetyltrimethlyammonium bromide (cetavlon). *Biochim. Biophys. Acta*. **10**, 607-612 (1953).

14. Meade, H.M., Long, S.R., Ruvkun, G.B., Brown, S.E. & Ausubel, F.M. Physical and genetic characterization of symbiotic and auxotrophic mutants of Rhizobium meliloti induced by transposon Tn5 mutagenesis. *J Bacteriol.* **149**, 114-122 (1982).

15. Gordon, D., Desmarais, C. & Green, P. Automated finishing with autofinish. *Genome Research.* **11**, 614-625 (2001).

16. Simpson, A.J. et al. The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature*. **406**, 151-157 (2000).